

## REMARKS/ARGUMENTS

The Examiner rejected claims 1 - 13 under §103(a), as being as being unpatentable over Kim et al. ("Acquisition of Semantic Patters for Information Extraction from Corpora")(hereinafter "Kim") and Soderland ("Learning Information Extraction Rules for Semi-structured and Free text")(hereinafter "Soderland"). As discussed more fully and completely below, neither reference discloses or suggests all of the elements of the claimed inventions for use in generating text analysis programs. Therefore, the combination of references cannot render obvious any of the pending claims.

In particular, the present invention teaches a method for automatically generating a multi-pass text analyzer using as input a hierarchy of text samples (i.e., user-designated units of interest or annotations culled from a set of input documents) to organize and dictate the manner in which said text analyzer will be constructed.

Kim does not teach a method for creating a text analyzer (such as a parser), but rather focuses on a method for generating *the semantic rules* of a semantic analyzer. Indeed, Kim requires a parser in order to supply the inputs for his system, while the present invention teaches methods for generating such a parser. Kim's methods can in no way be applied to the creation of the parser that he uses for his inputs.

The Examiner asserts that Soderland provides the hierarchy of text samples that is recited in the claimed inventions but absent in Kim. However, neither Kim nor Soderland make use of a hierarchy of text samples or annotations culled from a set of input documents. The fact that Soderland uses "semi-structured" input documents has no bearing on the patent-worthiness of the present invention. The claimed inventions use hierarchies of text samples or user-designated units of interest - structurally meaningful samples culled from input documents, such as "nouns," "noun phrases," and "adverbial phrases." Soderland makes use of "tagged texts" as part of the learning process, but those tags are not organized into a hierarchy or any other discernable representation. We note also that Kim makes use of entire sentences to find semantic events of interest within those sentences, rather than a hierarchy of text samples.

While ample precedent exists in the literature for generating single rule sets of various types, no precedent exists for automatically generating multi-pass analyzers in which each pass consists of a set of rules, and in which the multiple passes act repeatedly in a cascaded fashion to generate each subsequent pass of the analyzer. But this is the crux of the method that present invention teaches. We submit that the present invention teaches a novel and non-obvious method for automatically generating substantial portions of a complete multi-pass text analyzer.

We note that the review makes extensive use of Figure 2 in Kim, and makes assumptions and interpretations about that figure that are not corroborated by the methods discussed within the body of Kim. For example, the step of “merging into the knowledge base” is used to infer an entire method of generalizing, comparing, and merging rules, with no specific corroboration in the body of Kim.

Finally, we note that Kim teaches a method in which the user interacts with the system at various “decision points” to generate and accept semantic rules, while the present invention teaches a fully automated method for generating text analyzers, given a hierarchy of text annotations (i.e., “samples”) and attributes, the documents from which those annotations were culled, and the offsets of the annotations within their respective input documents. The roles of the user differ substantially in comparing the present invention and Kim’s methods.

We now respond to the specific discussion of claims in review.

### ***Claims 1 and 13***

Because claims 1 and 13 include similar elements (a)-(d), the following discussion applies to both claims. The Examiner rejected independent claims 1 and 13 contenting that Soderland teaches a hierarchy of text samples, anticipating claimed inventions, but such a hierarchy is nowhere to be found in Soderland. As defined in the specification, a “sample is a piece of text that users have decided is a unit of interest, such as a name or an idiomatic phrase.” The review contends that Soderland’s use of “semi-structured” input documents

anticipates a hierarchy of text samples (e.g., annotations) within the input documents, but the "semi-structured" nature of the input documents themselves in no way relates to the samples culled from those input documents. Thus, this feature of Soderland cannot serve as a basis for rejecting claim 1.

While the review "contends that this concept was well known in the art," the review provides no precedent for a hierarchy of text samples wherein those samples consist of annotations culled from a set of input documents. Because there is no evidence to support that such a precedent did exist, claim 1 is therefore valid.

Further, since Soderland uses no such hierarchy, Kim in view of Soderland cannot be said to use such a hierarchy, and the argument for rejection cannot be supported. Since the proposed combination fails to disclose or suggest at least claim element (a), it cannot render obvious claim 1 or 13, or any claims depending from those claims (i.e., claims 2-11). For at least these reasons, Applicant respectfully requests that the related rejections be withdrawn.

### ***Claim 2***

Because claim 2 depends from claim 1, the proposed combination cannot render claim 2 obvious for at least the reasons set forth above. Additionally, claim 2 is patentable on the following independent basis. With respect to claim 2, the Examiner states that Kim teaches "said rule is generalized into multiple rules and multiple passes." But Kim builds a semantic analyzer and not a text analyzer (of which a parser is a main embodiment). The semantic analyzer can be construed as having a single pass, and Kim makes no reference to generating multiple rule sets or multiple passes for a semantic analyzer. Nor can Kim's methods be used to generate a parser; rather, Kim's method relies on a pre-existing parser to supply needed inputs. In contrast, the claimed method focuses on the automatic generation of precisely the type of parser that Kim requires for input data.

The claimed method regarding multiple passes is not simply an artifice for separating a rule set into multiple parts. The text analyzer is constructed in such a fashion that each new pass constructed depends on the analysis performed by all the preceding passes. This

intricate interdependency is not disclosed or suggested by Kim, Soderland, nor indeed by any precedent to be found in the literature.

***Claim 3***

Because claim 3 depends from claim 1, it cannot be rendered obvious for at least the reasons set forth above. Additionally, claim 3 is patentable on the following independent basis. Kim does not teach or suggest that "multiple passes are added to said text analyzer." Kim's merging rules into a knowledge base serves to create a single set of rules, rather than multiple cascading rule sets (i.e., passes) as taught by the claimed invention.

***Claim 4***

Because claim 4 depends from claims 1 and 3, it cannot be rendered obvious for at least the reasons set forth above. Additionally, claim 4 is patentable on the following independent basis. Kim does not teach that "multiple passes are arranged in a cascading manner having a sequence of passes such that rules associated with a pass are applied to subsequent passes." The Examiner suggests that Kim teaches a method for generating a single set of semantic rules, which can be equated at most to a single pass of the text analyzer that the claimed method teaches. However, Kim does not even do this much, as he generates rules for a semantic analyzer, not a text analyzer, and uses a text analyzer (i.e., a parser) to produce inputs for generating a semantic analyzer. Nor can Kim's method generate such a parser, which is a typical output of the claimed method.

***Claim 5***

Because claim 5 depends from claim 1, it cannot be rendered obvious for at least the reasons set forth above. Additionally, claim 5 is patentable on the following independent basis. With respect to claim 5, Kim does not use text samples as defined in the application, nor does his method depend on the offsets of those samples within their respective text documents. Kim uses as input whole sentences as parsed by a pre-existing parser, with no need to refer to text offsets in generating semantic rules. Nor does Soderland use such hierarchies of text samples.

***Claim 6***

Because claim 6 depends from claims 1 and 4, it cannot be rendered obvious for at least the reasons set forth above. Additionally, claim 6 is patentable on the following independent basis. As we have described, the claimed invention teaches a fully automated method for generating rules and passes of a text analyzer. The user in the claimed method can set up a hierarchy of text samples (e.g., annotations) as well as other parameters for analyzer generation, after which the method automatically generates a text analyzer. Kim teaches a method in which the user is called upon during various parts of the semantic rule generation process, which is a substantially different process than that of the claimed invention.

Diverse methods for generating rules may involve a user at differing stages and for differing roles. The mere involvement of a user does not of itself imply an equivalence of methods.

***Claim 7***

Because claim 7 depends from claims 1 and 5, it cannot be rendered obvious for at least the reasons set forth above. Additionally, claim 7 is patentable on the following independent basis. As stated in Claim 7, the user in the claimed method performs distinct actions from the user in Kim's method. The user in the claimed method specifies general attributes used to modify the overall manner in which the rules and passes of a text analyzer are generated. The user in Kim's method, on the other hand, makes decisions regarding the generation and acceptability of specific semantic rules for a semantic analyzer, as those rules are presented by the system.

***Claim 8***

Because claim 8 depends from claims 5 and 1, it cannot be rendered obvious for at least the reasons set forth above. Additionally, claim 8 is patentable on the following independent basis. The claimed method relies on offsets in a parse tree in order to extract concepts used for generating syntactic rules of a text analyzer. As Kim's method deals with

inputs that already have been completely parsed by a pre-existing text analyzer, Kim's method does not extract concepts based on offsets in a parse tree. Kim's method attempts to map the parse tree for a sentence, already given as input, to pre-existing semantic "frames." This process relies on the identity of concepts in the parse tree, but not on their offsets. This process also fundamentally differs from the claimed method for generating syntactic (i.e., structural rules), which involves no such pre-existing semantic "frames," nor any equivalent representation.

#### ***Claim 9***

Because claim 9 depends from claims 6, 4 and 1, it cannot be rendered obvious for at least the reasons set forth above. Additionally, claim 9 is patentable on the following independent basis. As stated for Claims 6 and 7, the role of the user in the claimed methods and Kim's methods cannot be equated. In the claimed method, the user specifies attributes for the generation of passes and rules of a multi-pass text analyzer, while Kim's user makes decisions about the generation and acceptability of individual semantic rules as they are presented by the rule-generation method.

#### ***Claim 10***

Because claim 10 depends from claim 1, it cannot be rendered obvious for at least the reasons set forth above. Additionally, claim 10 is patentable on the following independent basis. While various automated rule generation methods involve the steps of creating, generalizing, and merging rules, the claimed methods for generating, generalizing, and merging rules cannot be equated with those of Kim. In the claimed invention, rules are compared based on the similarity of their syntactic structure and ordering. In the case of Kim, rules are compared to pre-existing semantic "frames" in order to determine if those rules should map to those frames.

The review cites Figure 2 of Kim in this and other cases to infer methods for generalization and merging of rules that are not corroborated elsewhere in the body of Kim. (p. 173, col. 2, par. 2) states "While merging, it is combined with other patterns to save space and save parsing time." This also provides insufficient information to infer in detail the

particular processes for generalization and merging are used in Kim, and cannot be the basis for rejecting specific claims and specific methods for such processes. Given that semantic rules are processed in Kim, while syntactic rules are processed in the claimed invention, and that Kim employs pre-existing "frames" while the claimed invention does not, these methods substantially differ at the outset.

***Claim 11***

Because claim 11 depends from claims 10 and 1, it cannot be rendered obvious for at least the reasons set forth above. Additionally, claim 11 is patentable on the following independent basis. We have already shown that the samples in the claimed inventions consist of annotations culled from input documents, while Kim's samples consist entirely of whole sentences, and Soderland uses a repository of texts. The claim limitation that such samples and their associated offsets and similar attributes may be represented in a file data structure is not equivalent to an assertion that the sentences of Kim may be stored in a file, nor to an assertion that Soderland's input documents consist of "semi-structured and free text."

In particular, the fact that Soderland uses input documents that are "semi-structured" does not correlate with the use of a hierarchy of text samples, wherein such samples are annotations culled from their input documents. Soderland's method uses hand-tagged texts, but the tagged portions of text are not organized into a hierarchy or any other representation apart from the input document containing those tags.

***Claim 12***

Applicant incorporates by reference, the arguments set forth above relative to claims 1 and 13. For at least these reasons, claim 12 is patentable over the proposed combination.

**CONCLUSIONS**

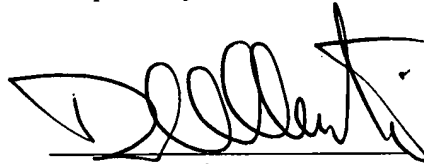
Applicant's invention is both novel and nonobvious over the prior art for the reasons set forth above. None of the prior art of record, either alone or in combination, teaches each and every element of Applicant's claimed invention.

Application No. 09/604,836  
Amdt. dated February 12, 2004  
Reply to Office action of November 14, 2003  
Attorney Docket Number 2100684-991110

For all of these reasons, Applicant respectfully asserts that all of claims 1-13 are in condition for allowance. The Examiner's early reconsideration is respectfully requested. If the Examiner has any questions, the Examiner is invited to contact Applicant's attorney at the following address or telephone number:

David Alberti  
c/o Patent Department  
GRAY CARY WARE & FREIDENRICH LLP  
2000 University Avenue  
East Palo Alto, CA 94303-2248  
Telephone: (650) 833-2052

Respectfully submitted,

A handwritten signature in black ink, appearing to read 'D. Alberti', with a stylized flourish at the end.

David Alberti  
Reg. No. 43,465

Dated: February 12, 2004